



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Nucleotide diversity in *Silene latifolia* autosomal and sex-linked genes

Citation for published version:

Qiu, S, Bergero, R, Forrest, A, Kaiser, VB & Charlesworth, D 2010, 'Nucleotide diversity in *Silene latifolia* autosomal and sex-linked genes' *Proceedings of the Royal Society B-Biological Sciences*, vol. 277, no. 1698, pp. 3283-3290. DOI: 10.1098/rspb.2010.0606

Digital Object Identifier (DOI):

[10.1098/rspb.2010.0606](https://doi.org/10.1098/rspb.2010.0606)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Proceedings of the Royal Society B-Biological Sciences

Publisher Rights Statement:

Freely available via Pub Med.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Nucleotide diversity in *Silene latifolia* autosomal and sex-linked genes

Suo Qiu^{1,2}, Roberta Bergero¹, Alan Forrest¹, Vera B. Kaiser¹
and Deborah Charlesworth^{1,*}

¹*Institute of Evolutionary Biology, School of Biological Sciences, The University of Edinburgh, Ashworth Laboratory, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK*

²*State Key Laboratory of Biocontrol and Key Laboratory of Gene Engineering of the Ministry of Education, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China*

The plant *Silene latifolia* has separate sexes and sex chromosomes, and is of interest for studying the early stages of sex chromosome evolution, especially the evolution of non-recombining regions on the Y chromosome. Hitch-hiking processes associated with ongoing genetic degeneration of the non-recombining Y chromosome are predicted to reduce Y-linked genes' effective population sizes, and *S. latifolia* Y-linked genes indeed have lower diversity than X-linked ones. We tested whether this represents a true diversity reduction on the Y, versus the alternative possibility, elevated diversity at X-linked genes, by collecting new data on nucleotide diversity for autosomal genes, which had previously been little studied. We find clear evidence that Y-linked genes have reduced diversity. However, another alternative explanation for a low Y effective size is a high variance in male reproductive success. Autosomal genes should then also have lower diversity than expected, relative to the X, but this is not found in our loci. Taking into account the higher mutation rate of Y-linked genes, their low sequence diversity indicates a strong effect of within-population hitch-hiking on the Y chromosome.

Keywords: *Silene latifolia*; nucleotide diversity; autosomal genes; sex chromosomes; genetic degeneration

1. INTRODUCTION

The plant *Silene latifolia* is of interest for studying the evolution of sex chromosomes from chromosomes of hermaphrodite ancestors (Bergero *et al.* 2007; Marais *et al.* 2008). Support for predictions that selection during the evolution of Y chromosomes will lower the effective population size (N_e) for Y-linked genes (Charlesworth & Charlesworth 2000), beyond the effect of the lower numbers of Y chromosomes than X chromosomes in populations, has been the finding that several Y-linked genes in *S. latifolia* have low diversity, compared with the homologous X-linked genes (Filatov *et al.* 2000, 2001; Ironside & Filatov 2005; Laporte *et al.* 2005). In non-recombining Y chromosome regions, selection will reduce N_e through hitch-hiking processes, including selective sweeps during the spread of advantageous alleles (causing selective sweeps that would lower diversity at linked silent sites, e.g. Rice 1987; Bachtrog 2004; Bachtrog *et al.* 2009), selective elimination of deleterious mutations (Charlesworth *et al.* 1993; Kaiser & Charlesworth 2009) and Muller's ratchet (Gordo & Charlesworth 2000; Gordo *et al.* 2002; Engelstädter 2008).

In evaluating the X–Y diversity difference as evidence for selection affecting Y chromosomes, other possible processes affecting diversity must be considered. The data so far collected cannot distinguish whether the difference is due to the predicted reduction in Y N_e , or to some process

increasing high species-wide diversity for X-linked genes. Possibilities include a higher X than Y mutation rate (although this is the opposite of the expectation for any difference between the mutation rates of the two chromosomes, and the data so far available support that expectation, see Filatov & Charlesworth 2002; Filatov 2005), sexual selection causing a high variance in male reproductive success (Laporte & Charlesworth 2002) or introgression of sequences from another species; hybridization is well known to occur between *S. latifolia* and *S. dioica* (Baker 1948; Minder & Widmer 2007) and seems to have led to sequences within each species having recombined into X-linked genes of the other (Atanassov *et al.* 2001; Laporte *et al.* 2005). Population size changes also affect the relative diversity levels of genome regions with different N_e (Fay & Wu 1999; Pool & Nielsen 2007).

A way to test whether the diversity is reduced for Y-linked genes, or elevated for X-linked ones, is to compare diversity with that at autosomal loci. A previous study of a single autosomal gene (*CCLS37.1*) found higher diversity than for a Y-linked gene (*SIY1*), but considerably lower than for an X-linked gene studied in the same set of plants, leaving the possibility open that X-linked genes might have unusually high diversity; however, this gene may have experienced a recent selective sweep (Filatov *et al.* 2001). Sequences are available for only one other *S. latifolia* autosomal gene, *Slop*, whose silent site diversity is 0.65 per cent (Muir & Filatov 2007). A third estimate, from an *Ap3* gene (Matsunaga *et al.* 2003), is no longer relevant, because the gene is probably X-linked (Ishii *et al.* 2008).

We therefore obtained new data on DNA sequence diversity in *S. latifolia* autosomal loci, and also increased

* Author for correspondence (deborah.charlesworth@ed.ac.uk).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2010.0606> or via <http://rspb.royalsocietypublishing.org>.

Table 1. Genes studied in this paper and the primers used for PCR amplification.

gene	best blast hit to <i>A. thaliana</i> genome		primers	regions studied	length (bp)
	locus name	accession number			
<i>E72</i>	pentatricopeptide repeat-containing protein	AT1G02150	E72-f: ATTTAATTGCTAAAGTTCATGGAATTCC E72-r: CATTAGAAGGTTTCCGATTTTAGGGTC	exon2	674
<i>E106</i>	zinc ion-binding protein (ATVPS11)	AT2G05170	E106-f: GTGTCATTTTCAGATCAAAGAAGTGG E106-r: ATACTCCAAGGCTTCATCATATCTACC	exon4	656
<i>E163</i>	vacuolar ATP synthase subunit D (VATD)	AT3G58730	E163-f: ACGCAACGCCTTAATGTGGTAC E163-r: CCTACGGTTTGTAGTCTTGATAGC	exon1	410
<i>E265</i>	FCLY (farnesylcysteine lyase) or prenylcysteine oxidase	AT5G63910	E265-f: GGACATGCTTAAATGGTCTGGTTTATATG E265-r: AAAGGTGTAGCAAAAACCGTCACT TG	intron2	464
<i>E304</i>	putative transaldolase	AT5G13420	E304-f: CTTATCAAATCCGGAGTAGACATTG E304-r: CAAGCATCTTATCAACCAGGGAATC	exon4 intron4	244 392–397
<i>E314</i>	rhodanese-like domain-containing protein	AT1G34000	E314-f: AACATCAGTACCCTGCATGAGAATG E314-r: TGGAGAGGATGAGTTTTAGTTGATC	exon1	344
<i>E391</i>	proton motive force-dependent protein transmembrane transporter (HCF106)	AT5G54770	E391-f2: CGCAAACCAGCACATCTATTTCTAG E391-r2: GTTGAGCAATAAGTGAGATGGGGAC	intron1 exon2	253–307 413
<i>E393</i>	photosystem I subunit L (PSAL)	AT4G12800	E393-f: CAACTAGGGCTTTATTTGTCACACC E393-r: TACAAGAGAAAGTAAGCCCAAACAAC	intron1 exon2	13 419

the sample size for sex-linked genes by adding diversity data for three further gene pairs with homologues on the X and Y chromosomes, *SlCyp* and *Sl7* (Bergero *et al.* 2007), and a newly discovered gene, *SlXY9* (Kaiser *et al.* in press). Our new data, with more loci, and also from more *S. latifolia* populations than previous studies, show that Y diversity is considerably reduced, supporting the ongoing genetic degeneration hypothesis.

2. MATERIAL AND METHODS

(a) *Genes studied*

To obtain autosomal genes, a large number of EST sequences were obtained by 454 sequencing of *S. latifolia* cDNAs (R. Bergero, S. Qiu, H. Borthwick, A. Forrest & D. Charlesworth 2008–2010, unpublished data). The *Arabidopsis thaliana* genome was used to identify likely orthologues, and genes belonging to large gene families (in *A. thaliana*) were discarded, as these are unlikely to provide simple genetic segregation data. Translated cDNA sequences were aligned with *A. thaliana* coding sequences (Swarbreck *et al.* 2008), using the software FASTA (Pearson 1990). Single genes with e -value $< 10^{-5}$ were retained and annotated in Sequencher v.4.5 (GeneCodes, Ann Arbor, MI, USA), using information from *A. thaliana* orthologues to suggest intron positions and functions. The genes studied, and the polymerase chain reaction (PCR) primers for each gene are given in table 1.

(b) *Silene latifolia* families and population samples

To identify autosomal (and pseudoautosomal) genes, we studied segregation of length or sequence variants detected

in families (following the approach previously used in this species, see Bergero *et al.* 2007). To study diversity, we sequenced eight autosomal genes, using a set of *S. latifolia* male plants (electronic supplementary material, table S2) from across Europe, including much of the likely native distribution of this species. These plants were grown from seeds collected from naturally pollinated females, and one male was sampled per locality, to provide a sample suitable for estimating effective population sizes from nucleotide diversity, and in which variant frequencies should follow a neutral coalescent unless selection occurs (whereas samples with multiple individuals per population are expected to have a different variant frequency spectrum, see Wakeley & Aliacar 2001). We also estimated diversity for two recently characterized sex-linked genes, *SlCyp* and *Sl9X*, which have been shown to be single-copy genes (Bergero *et al.* 2007), using two males per locality to obtain similar numbers of X- and Y-linked alleles to the autosomal sample. Because our samples from individual populations are small, our analyses of subdivision (see below) divided the *S. latifolia* populations into four geographical regions (electronic supplementary material, table S2), roughly corresponding to previously suggested races (e.g. Mastenbroek & Van Brederode 1986).

The orthologous genes were also sequenced from some plants from the closely related dioecious species, *S. dioica* (electronic supplementary material, table S2), and from *S. vulgaris*, a related species that does not have sex chromosomes, but whose populations are gynodioecious (with females and hermaphrodites, see Marsden-Jones & Turrill 1957; Baker 1966; Charlesworth & Laporte 1998).

(c) DNA extraction, PCR reaction and cloning

For the segregation study, seeds were germinated on Murashige and Skoog (Murashige & Skoog 1962) medium. DNA was extracted from 20-day-old seedlings using the FastDNA kit (Q-biogene) following the manufacturer's instructions. Seedlings' sexes were determined using a 293-bp indel fixed in intron 2 of all Y-linked copies of the *SlCyp* gene (Bergero et al. 2007). For the sequence diversity study, DNA was extracted from leaves collected from mature plants grown in greenhouse conditions, using the same method.

Primers for the genes newly sequenced in our study were designed using the EST sequences of *S. latifolia* (table 1). The primers for *SlXY7* were SIX7f3 (GAATGGCA-GAAATGTGGCAATGCAAC) and SIX7r2 (GGAAA GCTCCTTTTCAACGAAGTC), and, for *SlCypXY*, cypXf3 (AATTTTCTGTACATCGCCTGATCGCAG) and cypR1 (AGTGGATCTCCTGTCTGTATCATGAAACC). Amplifications used a Finnzymes' Piko Thermal Cycler with Phire hot-start DNA polymerase (Finnzyme), under the following PCR conditions: one cycle of initial denaturation at 98°C for 30 s, 10 cycles at 98°C for 5 s, 65°C for 5 s, and 72°C for 50 s, and then 25 cycles of 98°C for 5 s, 60°C for 5 s and 72°C for 1 min, and finally one cycle at 72°C for 5 min.

The PCR amplicons were cleaned with ExoSAP-IT (Amersham Biosciences, Tokyo) and sequenced on an ABI 3730 capillary sequencer (Applied Biosystems). When no indel variants were present, the single nucleotide polymorphisms (SNPs) were verified by direct sequencing from both strands. Most such variants were found in multiple individuals. Amplicons containing intron regions were sometimes heterozygous for indels. In such individuals, and for plants with singleton variants, the DNA was re-amplified with a proofreading DNA polymerase (Phusion, Finnzyme) using the PCR conditions given above, and cloned into a T-tailed pBSKS+ vector (Stratagene) before sequencing.

As described below, two autosomal genes have high diversity. Our genetic results suggest that they are single-copy genes. However, these data apply only to the parents of the families used, and furthermore the genetic results for *E393* are based on a single SNP, while for *E391* we used an intron region, not the region whose diversity was estimated. To test for the existence of paralogous copies of these genes among the plants used in the diversity survey, we sequenced plants in family H2005-1 (electronic supplementary material, table S1), with the primers used in the survey. For *E393*, both the F1 parental plants are heterozygous for 13 SNPs, and 10 progeny sequenced include heterozygotes and both homozygotes, as expected for segregation of a single gene. For *E391*, the parents are heterozygous for a single SNP, which also segregated in the progeny. Therefore, in this family, these genes do not appear to be duplicated.

(d) Sequence analyses

The newly obtained sequences of each gene were first aligned with Sequencher v.4.5 and manually adjusted using the sequence alignment editor Se-Al v.2.0a11 (Se-Al: Sequence Alignment Editor, <http://evolve.zoo.ox.ac.uk/>). The alignments were submitted to GenBank, under accession numbers HM188567–HM189181. Polymorphism and haplotype analyses (including estimates of nucleotide diversity and population subdivision) were done using DNASP v.5.00.06 (Librador & Rozas 2009) and MEGA 4 (Tamura et al. 2007). We estimated diversity for silent sites, because this relates closely to N_e values (Loewe et al. 2006).

Hudson-Kreitman-Aguadé (HKA) tests for differences in diversity between sets of genes (see §3) were done with a maximum-likelihood implementation of the test, multi-locus HKA (mlHKA) (Wright & Charlesworth 2004). These tests used the *S. vulgaris* sequences mentioned above, and divergence from the *S. latifolia* orthologues was estimated using DNASP. Silent site divergence values for these two species range from about 9 to 22 per cent. We also did Tajima's (Tajima 1989a) and Fay and Wu's tests for selection, which depend on variant frequencies (see §3), and the DH test, which is designed to be less sensitive to changes in recent demographic history (Zeng et al. 2006). For these tests, we used variants of all site types, and *S. vulgaris* as the outgroup.

In addition, the diversity data were compared with published estimates from four other sex-linked genes, *SlXY1* (Filatov et al. 2001) and *SlXY4* and *SIDD44* (Ironside & Filatov 2005; Laporte et al. 2005), which used smaller samples of European *S. latifolia*, and *SlSs* (Filatov 2008), which used samples similar to those used here.

We tested for mutation rate differences in four new X–Y gene pairs, three genes used in our diversity analyses (*SlXY7*, *Cyp-XY*, *SlXY9*) and *XY6a*, which was not included in the diversity survey, because a duplicate gene exists (Bergero et al. 2007). We extracted non-coding and fourfold degenerate sites using DNASP v.5.00.06 (Librador & Rozas 2009), and used the BASEML program of the PAML v.3e package (Yang 2001) to obtain maximum-likelihood estimates of the relative substitution rates and perform likelihood ratio tests of the significance of silent substitution rate differences. We estimated the likelihood under two models (i) with three evolutionary rate parameters, one estimated for each of the X-linked, Y-linked and non-sex-linked (*S. vulgaris*), branches, and (ii) with the same rate for X- and Y-linked genes. Significance of differences between the X- and Y-linked genes was tested using a χ^2 distribution with 1 d.f.

3. RESULTS

(a) Discovery of autosomal genes

The eight autosomal genes in our diversity survey (table 1) were found to be single-copy and not sex-linked by studying their segregation ratios in families. Electronic supplementary material, table S1 shows that these genes almost always show the expected ratios, with no heterogeneity in ratios between the two sexes of progeny (other than for locus *E163*, which deviated significantly from the expected ratio in the female progeny, but with no significant heterogeneity between the sexes). Two further genes, *E284* and *E241*, appear to be pseudoautosomal, and will be described in detail elsewhere.

Six of the eight autosomal genes also map to autosomal linkage groups with multiple other genes (R. Bergero, A. Forrest, H. Borthwick, S. Qiu & D. Charlesworth, unpublished data). For two genes, *E163* and *E265*, pseudo-autosomal locations are not ruled out from these data. However, the genotypes in family G2008-3 (electronic supplementary material, table S1) allowed us to test *E265* for linkage to the pseudoautosomal gene *E241*; no linkage was detected.

The haplotypes from males sampled from the natural populations further support autosomal inheritance for

Table 2. Diversity statistics for *S. latifolia* autosomal genes.

gene	<i>n</i> ^a	sites in alignment	number of polymorphic sites ^b	number of different haplotypes ^c	number of homozygotes
<i>E72</i>	46	674	39	44	1
<i>E106</i>	48	656	6	7	15
<i>E163</i>	48	410	9	9	14
<i>E265</i>	48	674	5	8	16
<i>E304</i>	47	676	44	33	8
<i>E314</i>	46	574	40	38	2
<i>E391</i>	48	413	27	36	11
<i>E393</i>	48	432	38	34	8

^aNumber of alleles sequenced.
^bExcluding indels.
^cIncluding indels.

all eight genes. If a locus is sex-linked, we expect variants differentiating the Y-linked alleles from all X sequences. Thus males should always be heterozygous (unless the Y chromosome has lost the gene, in which case all males will appear to be homozygotes). Finding both heterozygous and homozygous males, as we found for all loci, including *E163* and *E265* (table 2) suggests that a gene is in a recombining region of the genome (it is unlikely that they carry an allele that does not amplify with our primers, because no null homozygotes were seen for any locus in our study).

(b) Autosomal locus diversity results and comparisons with X and Y

Figure 1 summarizes the currently available nucleotide diversity estimates based on silent sites (π_{silent}) for all genes studied so far, including previously published results, and three newly sequenced sex-linked genes, *SLXY7*, *SLXY9* and *SLCypXY*. The Y-linked genes have much lower diversity values than X-linked or autosomal genes. The X–Y difference is significant by a multi-locus HKA test, using divergence from the homologous *S. vulgaris* sequences ($p = 1.06 \times 10^{-5}$), excluding *SLsX*, where no Y sequences are available, and either including or excluding *SLAp3* (where X-linkage is not firmly established, see Ishii *et al.* 2008); the Y-autosomal significance value was $p = 2.35 \times 10^{-4}$. Overall, π_{silent} for the eight autosomal loci is 2.88, versus 2.15 per cent for the seven X-linked genes, and 0.14 per cent for their Y-linked homologues. Although the diversity estimates vary widely among loci, the X-autosomal difference is significant ($p = 0.0017$) based on all the X-linked genes, but not if we exclude *SLsX* ($p = 0.92$), which may have undergone a selective sweep (Filatov 2008), and differs significantly from the other loci.

The overall *X/A* diversity ratio is 0.747, very close to the value (0.75) expected with a 1 : 1 sex ratio and the same variance in male and female reproductive success, so that the effective size for X-linked genes is 0.75 that of autosomal genes (e.g. Laporte & Charlesworth 2002; Keinan *et al.* 2009). These assumptions predict an effective size for Y-linked genes one-fourth that of autosomal genes, but the *Y/A* silent site diversity ratio is 0.049 (one-twentieth). Together, these results suggest hitch-hiking processes caused by selection on the non-recombining Y chromosome, lowering the effective size for Y-linked genes below the expected value (see §1).

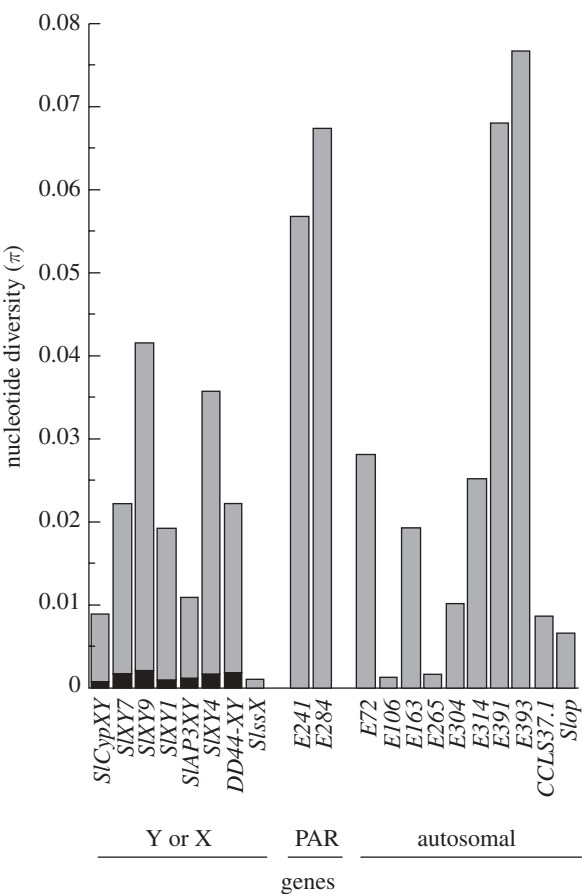


Figure 1. Summary of nucleotide diversity estimates using silent sites for X, A and PAR genes of *S. latifolia* (grey bars) and for the Y-linked homologues of the X-linked loci (black bars). The numbers of sites analysed, and references for the data sources are in the electronic supplementary material, table S3.

Two autosomal genes (*E393* and *E391*) have very high diversity values, but mlHKA tests within the set of autosomal genes detect no statistically significant differences for these genes (see below). Our genetic results suggest that the *E393* and *E391* genes do not have duplicate copies, and neighbour-joining trees of the sequences contain no separate clusters of *S. latifolia* sequences (electronic supplementary material, figure S1). The high diversity values are not due to high mutation rates for these genes, because divergence estimates from

Table 3. Tests for selection for autosomal and X-linked genes, using all site types. Statistically significant values ($p < 0.05$) are shown in bold.

	Tajima's test		Fay and Wu's test		DH test value <i>p</i> -value
gene	<i>D</i> _T	<i>p</i> -value	<i>H</i>	<i>p</i> -value	
autosomal genes					
<i>E72</i>	−0.941	0.179	0.373	0.507	0.293
<i>E106</i>	−1.44	0.058	0.392	0.510	0.245
<i>E163</i>	−0.167	0.499	−0.601	0.167	0.252
<i>E265</i>	−0.654	0.300	−1.99	0.0416	0.0872
<i>E304</i>	−1.77	0.0158	−1.353	0.0761	0.0268
<i>E314</i>	−0.051	0.552	−1.885	0.0424	0.330
<i>E391</i>	0.070	0.6021	−0.762	0.143	0.380
<i>E393</i>	0.305	0.693	−1.088	0.102	0.488
X-linked genes					
<i>SlCypX</i>	−1.196	0.105	0.258	0.437	0.233
<i>SlX7</i>	−1.362	0.0775	−3.093	0.0154	0.0146
<i>Sl9X</i>	0.320	0.695	−0.480	0.192	0.487
<i>SlX1</i>	−0.318	0.425	0.183	0.399	0.246
<i>SLAP3A(X)</i>	−1.705	0.0354	0.807	0.821	0.659
<i>SlX4</i>	−0.258	0.461	−1.571	0.0588	0.268
<i>DD44-X</i>	0.033	0.553	0.470	0.559	0.357
<i>SlSsX</i>	−1.94	0.007	−1.722	0.050	0.016

S. vulgaris and *S. dioica* sequences are similar to the values for other loci. Finally, we tested for introgression in our *S. latifolia* populations, using sequences from *S. dioica* sampled from Southern Finland, Scotland, England, France and Italy (electronic supplementary material, table S2 and figure 1; both species co-occur in all regions except Finland). We found no fixed differences between the species (precluding a formal analysis of introgression), and, even with our non-exhaustive sampling, many polymorphisms are shared (24 for *E393* and 7 for *E391*, versus 18 and 12 polymorphisms, respectively, exclusive to *S. latifolia*). *Silene dioica* sequences are also diverse (silent site $\pi > 5\%$ for both loci), but the haplotypes are mostly distinct from those of *S. latifolia* plants. The few clusters with sequences from both species have very low bootstrap support, and the sequences are mostly not extremely similar (electronic supplementary material, figure S1); for *E391*, no *S. latifolia* sequences are among the main *S. dioica* cluster. For *E393*, excluding two *S. latifolia* alleles that cluster with *S. dioica* sequences reduces diversity only slightly (from 0.077 to 0.075). Introgression from *S. dioica* thus seems unlikely to be the sole reason for high diversity of these genes, but we cannot exclude past introgression from other species not currently sympatric with *S. latifolia*.

We also tested for a higher Y than X mutation rate in the four new X–Y genes. The ratio of Y- to X-linked substitution rates, $R_{Y/X}$, exceeds 1 for all four genes, two of them significantly (*SlCyp*, with $p = 0.039$ and $R_{Y/X} = 2.03$, and *SlXY6a*, with $p < 0.001$ and $R_{Y/X} = 2.80$), while two short sequences (approx. 100 bp analysed) were non-significant (*SlXY7*, with $R_{Y/X} = 1.08$, and *SlXY9*, with $R_{Y/X} = 1.42$).

(c) Tests for selection

The overall X/A ratio of diversity estimates is close to the expected value of 0.75. However, diversity varies among

both the X-linked and the autosomal genes. The X/A ratio would be affected if some genes in either type of location had been subject to recent selective sweeps, or have high diversity owing to long-term balancing selection, and it is thus possible that the agreement is coincidental. As well as the two autosomal genes with high silent site diversity discussed above, two (*E106* and *E265*) have diversity values even lower than *CCLS37.1* (which, as mentioned above, may have undergone a selective sweep). The recent selective sweep at the X-linked *SlSsX* gene (Filatov 2008) is supported by mlHKA tests among the X-linked loci (see above). Only the two low-diversity autosomal genes (*E106* and *E265*) differ significantly from the other autosomal loci by mlHKA tests. These may thus have undergone recent selective sweeps, though the test is highly significant only for *E265* ($p = 2 \times 10^{-5}$; for *E106*, $p = 0.026$).

We also detect an unexpectedly high proportion of low-frequency variants, compared with the expectation at equilibrium under neutrality. Tajima's D -values for autosomal genes (using either silent sites or all site types) are mostly negative, often strongly so, but only *E304* has an individually significantly negative value (table 3). This overall pattern could be caused by an extreme bottleneck (Tajima 1989a,b), but is more likely to be due to population expansion (Tajima 1993; Innan & Stephan 2000; Ray et al. 2003), which is plausible for this species, which probably expanded after the last ice age. Subdivision may also contribute, since sampling few individuals from many localities can produce negative D -values (Ptak & Przeworski 2002). However, D becomes even more negative when we use a single sequence per population, which produces a standard coalescent process if geographical structure is unimportant (Wakeley & Aliacar 2001; De & Durrett 2007). Geographical structure often produces positive Tajima's D -values, owing to local adaptation, but sampling mostly from a common subpopulation or race,

with few individuals from a different race, could produce an unexpected frequency of rare variants. However, the clustering program STRUCTURE2 (Pritchard *et al.* 2000) identifies no geographical structure in *S. latifolia*, using six X-linked genes (G. Muir, R. Bergero, D. Charlesworth & D. A. Filatov 2007–2010, unpublished data).

We also applied the DH test for selection, which is less sensitive than other tests to changes in populations' recent demographic history (Zeng *et al.* 2006). The *E304* gene again yielded a significant test result, and two other genes gave marginally significant Fay and Wu's tests (table 3). None of the tests were significant for any Y-linked gene (results not shown). Of the five X-linked genes, selection at the *S/SsX* locus is confirmed, and is inferred also for *S/X7* (table 3). Overall, consistent with the HKA test results described above, there is clearly no convincing evidence for selection in or near most of the loci studied, and thus for excluding loci in estimating the X/A ratio; without the low diversity *S/SsX* gene, this ratio increases to 0.866, but excluding *E265* also, it becomes 0.688.

Finally, we applied haplotype tests, to test for balancing selection (for the two loci with high diversity, *E391* and *E393*, such tests might also detect evidence for paralogous copies). We used DNASP to run neutral coalescent simulations, using the estimated recombination rate for each locus, and obtain 95% confidence intervals for the expected numbers of haplotypes. The observed values for all loci were within these intervals.

(d) Population subdivision and comparisons between the X and Y

Another possible explanation for high silent site diversity values is population subdivision, but this seems not to be the cause. We analysed diversity separately for the samples from individual populations or from four geographical regions (see §2 and electronic supplementary material, table S2). Despite the large distances between the locations where the plants were sampled, K_{ST} values for the autosomal genes (electronic supplementary material, table S4) are not high (the average is 0.125 for silent sites, similar to the value for human populations), with similar values for the different loci. For the sex-linked genes, K_{ST} is higher (the X and Y averages are 0.177 and 0.245, reflecting the fact that they both have lower N_e). Estimates using individual localities are higher (electronic supplementary material, table S4), but our sample size of two alleles per locality must considerably overestimate subdivision, because such samples often contain only the commonest allele, and monomorphism elevates K_{ST} values (Charlesworth 1998).

4. CONCLUSIONS AND DISCUSSION

Our results establish that diversity is low for Y-linked genes, rather than being unusually high for X-linked ones, since the mean X diversity is slightly lower than the autosomal loci, and close to the value expected with a 1:1 sex ratio and the same variance in male and female reproductive success. Hitch-hiking processes on the non-recombining Y chromosome are predicted to lower Y diversity, but a possible alternative reason for low diversity of Y-linked genes is a high variance in

male reproductive success. The autosomes should then also have lower diversity than expected, relative to the X, i.e. the X/A ratio should be increased above 0.75 (Laporte & Charlesworth 2002; Ellegren 2009). We therefore conclude that there is no evidence that sexual selection is important in populations of this plant.

Another possibility, however, is a recent change in the demographic history of the population. Changes in population size affect the X and Y chromosomes differently from the autosomes, and can increase or reduce X/A and Y/A diversity ratios for many generations (Fay & Wu 1999; Pool & Nielsen 2007). Rapid growth in population size leads to increased X/A ratios, lasting for considerable times after the size change, whereas a reduction in population size decreases the ratio (the decrease can be severe, but not for long, see Pool & Nielsen 2007). A temporary bottleneck generally lowers X/A ratios, but a very severe bottleneck, followed by growth as the population recovers, can have the opposite effect. The Y/A diversity ratio (and the mitochondrial/autosomal one) are affected in similar ways, but the low effective size of Y-linked genes causes their diversity to recover much faster than for X-linked genes, so that prolonged effects of a bottleneck are less likely to be found.

We therefore compared diversity estimates from two geographical regions, since the size change models predict different ratios in populations whose sizes differ. Although the regions are not strongly isolated (see above), the samples from northern populations (as defined in §2), when analysed separately, tend to have slightly higher diversity than southern ones (14 of 17 genes for which data are currently available), with a south/north diversity ratio for silent sites of 0.78 for autosomal genes (but no difference for the Y-linked sequences). The difference is more pronounced for the X-linked genes: southern populations' silent site diversity is only 56 per cent of that in the northern sample, so that their X/A ratio is much lower (0.35) than the overall value, consistent with a greatly reduced southern population in the recent past. However, Tajima's *D*-values are not more negative in the southern sample.

A low population size in the recent past, from which the population has not yet recovered, could temporarily greatly reduce diversity for Y-linked genes, relative to X-linked ones, and is thus a possible alternative to hitch-hiking to explain the observed low Y sequence diversity. Figure 2 illustrates this. A larger decrease in size could give an even larger X/Y ratio. On the other hand, any such difference will be lessened if Y genes have a higher mutation rate than X ones (as seems to be the case in *S. latifolia*, see above). It therefore seems unlikely that a population size change can explain the ratios observed in *S. latifolia*.

Our observed higher K_{ST} values for Y-linked genes than for other loci, regardless of the spatial scale of the samples used, suggest that the reduction in Y diversity is mainly caused by within-population processes, such as elimination of deleterious mutations, or local adaptations involving Y-linked genes, and is not principally an effect of a species-wide selective sweep owing to an advantageous mutation fixing in a Y-linked gene (which should lead to lower K_{ST} for the Y, until equilibrium is re-established). The highest within-population X/Y diversity ratios average 31 for silent sites (electronic supplementary material,

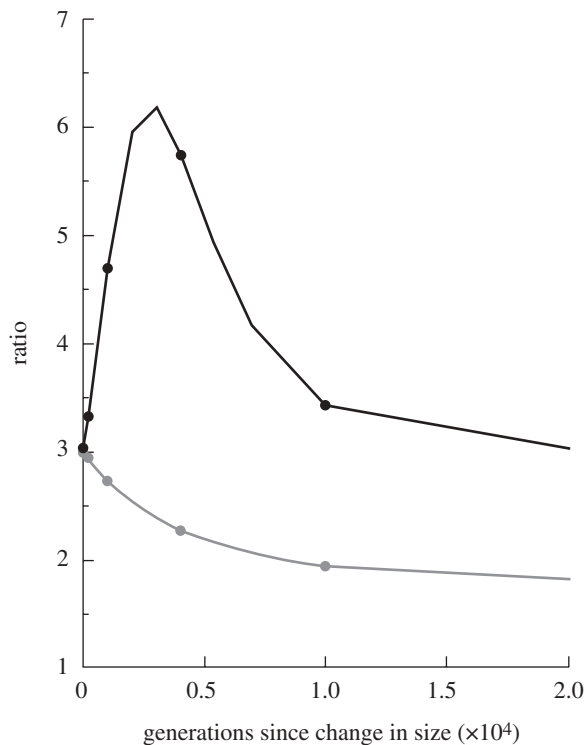


Figure 2. Plot X/Y ratios predicted using equation (4) of Pool and Nielsen (Pool & Nielsen 2007). The calculations assumed the same mutation rate for all chromosomes, and a fivefold decrease (black symbols and line) or increase (grey symbols and line) in population size from an initial value of 10 000.

table S4). Overall, we therefore estimate that *S. latifolia* Y-linked sequences have a 20–30-fold reduction in effective population size, consistent with the previous estimates using slightly larger within-population sample sizes, but fewer localities and loci (Laporte *et al.* 2005). Taking into account the higher mutation rate of Y-linked genes, the low diversity of Y-linked genes is quantitatively consistent with strong within-population hitch-hiking processes associated with the ongoing Y chromosome genetic degeneration (Kaiser & Charlesworth 2009). However, the different kinds of processes are not exclusive, and a species-wide selective sweep could also have contributed to reducing diversity.

We are grateful to NERC for 454 sequencing, to BBSRC for funding, Kai Zeng for help with DH tests and A. Betancourt for help with PAML analyses. S.Q. was supported by 973 Programme of China (2007CB815701) and the Chang Hung-Ta Science Foundation of Sun Yat-sen University.

REFERENCES

Atanassov, I., Delichère, C., Filatov, D. A., Charlesworth, D., Negrutiu, I. & Monéger, F. 2001 A putative monofunctional fructose-2,6-bisphosphatase gene has functional copies located on the X and Y sex chromosomes in white campion (*Silene latifolia*). *Mol. Biol. Evol.* **18**, 2162–2168.

Bachtrog, D. 2004 Evidence that positive selection drives Y-chromosome degeneration in *Drosophila miranda*. *Nat. Genet.* **36**, 518–522. (doi:10.1038/ng1347)

Bachtrog, D., Jensen, J. & Zhang, Z. 2009 Accelerated adaptive evolution on a newly formed X chromosome. *PLoS Biol.* **7**, e1000082.

Baker, H. G. 1948 Stages in invasion and replacement demonstrated by species of *Melandrium*. *J. Ecol.* **36**, 96–119.

Baker, H. G. 1966 The evolution of floral heteromorphism and gynodioecism in *Silene maritima*. *Heredity* **21**, 689–692. (doi:10.1038/hdy.1966.67)

Bergero, R., Forrest, A., Kamau, E. & Charlesworth, D. 2007 Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics* **175**, 1945–1954. (doi:10.1534/genetics.106.070110)

Charlesworth, B. 1998 Measures of divergence between populations and the effect of forces that reduce variability. *Mol. Biol. Evol.* **15**, 538–543.

Charlesworth, B. & Charlesworth, D. 2000 The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572. (doi:10.1098/rstb.2000.0717)

Charlesworth, D. & Laporte, V. 1998 The male-sterility polymorphism of *Silene vulgaris*. Analysis of genetic data from two populations, and comparison with *Thymus vulgaris*. *Genetics* **150**, 1267–1282.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

De, A. & Durrett, R. 2007 Stepping-stone spatial structure causes slow decay of linkage disequilibrium and shifts the site frequency spectrum. *Genetics* **176**, 969–981. (doi:10.1534/genetics.107.071464)

Ellegren, H. 2009 The different levels of genetic diversity in sex chromosomes and autosomes. *Trends Genet.* **25**, 278–284. (doi:10.1016/j.tig.2009.04.005)

Engelstädter, J. 2008 Muller's ratchet and the degeneration of Y chromosomes: a simulation study. *Genetics* **180**, 957–967. (doi:10.1534/genetics.108.092379)

Fay, J. C. & Wu, C. I. 1999 A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol. Biol. Evol.* **16**, 1003–1005.

Filatov, D. A. 2005 Substitution rates in a new *Silene latifolia* sex-linked gene, *SlssX/Y*. *Mol. Biol. Evol.* **22**, 402–408. (doi:10.1093/molbev/msi003)

Filatov, D. 2008 A selective sweep in or near the *Silene latifolia* X-linked gene *SlssX*. *Genet. Res.* **90**, 85–95.

Filatov, D. A. & Charlesworth, D. 2002 Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *S. dioica*. *Mol. Biol. Evol.* **19**, 898–907.

Filatov, D. A., Monéger, F., Negrutiu, I. & Charlesworth, D. 2000 Evolution of a plant Y-chromosome: variability in a Y-linked gene of *Silene latifolia*. *Nature* **404**, 388–390. (doi:10.1038/35006057)

Filatov, D. A., Laporte, V., Vitte, C. & Charlesworth, D. 2001 DNA diversity in sex linked and autosomal genes of the plant species *Silene latifolia* and *S. dioica*. *Mol. Biol. Evol.* **18**, 1442–1454.

Gordo, I. & Charlesworth, B. 2000 The speed of Muller's ratchet with background selection, and the degeneration of Y chromosomes. *Genet. Res.* **78**, 149–162.

Gordo, I., Navarro, A. & Charlesworth, B. 2002 Muller's ratchet and the pattern of variation at a neutral locus. *Genetics* **161**, 835–848.

Innan, H. & Stephan, W. 2000 The coalescent in an exponentially growing metapopulation and its application to *Arabidopsis thaliana*. *Genetics* **155**, 2015–2019.

Ironside, J. E. & Filatov, D. A. 2005 Extreme population structure and high intraspecific divergence of the *Silene latifolia* Y chromosome. *Genetics* **171**, 705–713. (doi:10.1534/genetics.105.041210)

Ishii, K., Fujita, N., Nishiyama, R. & Kawano, S. 2008 Alleles of the sex chromosomes and their evolution in *Silene latifolia*. *Genes Genet. Syst.* **83**, 511.

- Kaiser, V. B. & Charlesworth, B. 2009 The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet.* **25**, 9–12. (doi:10.1016/j.tig.2008.10.009)
- Kaiser, V. B., Bergero, R. & Charlesworth, D. In press. A new plant sex-linked gene with high sequence diversity and possible introgression of the X copy. *Heredity*
- Keinan, A., Mullikin, J., Patterson, N. & Reich, D. 2009 Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nat. Genet.* **41**, 66–70. (doi:10.1038/ng.303)
- Laporte, V. & Charlesworth, B. 2002 Effective population size and population subdivision in demographically structured populations. *Genetics* **162**, 501–519.
- Laporte, V., Filatov, D. A., Kamau, E. & Charlesworth, D. 2005 Indirect evidence from DNA sequence diversity for genetic degeneration of Y-chromosome in dioecious species of the plant *Silene*: the SIY4/SIX4 and DD44-X/DD44-Y gene pairs. *J. Evol. Biol.* **18**, 337–347. (doi:10.1111/j.1420-9101.2004.00833.x)
- Librador, P. & Rozas, J. 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.
- Loewe, L., Charlesworth, B., Bartolomé, C. & Noël, V. 2006 Estimating selection on nonsynonymous mutations. *Genetics* **172**, 1079–1092. (doi:10.1534/genetics.105.047217)
- Marais, G., Nicolas, M., Bergero, R., Chambrier, P., Kejnovsky, E., Monéger, F., Hobza, R., Widmer, A. & Charlesworth, D. 2008 Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *Curr. Biol.* **18**, 545–549. (doi:10.1016/j.cub.2008.03.023)
- Marsden-Jones, E. M. & Turrill, W. B. 1957 *The bladder champions*. London, UK: Ray Society.
- Mastenbroek, O. & Van Brederode, J. 1986 The possible evolution of *Silene pratensis* as deduced from present day variation patterns. *Biochem. Syst. Ecol.* **14**, 165–181. (doi:10.1016/0305-1978(86)90058-X)
- Matsunaga, S., Isono, E., Kejnovsky, E., Vyskot, B., Kawano, S. & Charlesworth, D. 2003 Duplicative transfer of a MADS box gene to a plant Y chromosome. *Mol. Biol. Evol.* **20**, 1062–1069. (doi:10.1093/molbev/msg114)
- Minder, A. & Widmer, A. 2007 A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Mol. Ecol.* **17**, 1552–1563. (doi:10.1111/j.1365-294X.2008.03709.x)
- Muir, G. & Filatov, D. 2007 A selective sweep in the chloroplast DNA of dioecious *Silene* (Section *Elisanthe*). *Genetics* **177**, 1239–1247. (doi:10.1534/genetics.107.071969)
- Murashige, T. & Skoog, F. 1962 A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant* **15**, 473–497. (doi:10.1111/j.1399-3054.1962.tb08052.x)
- Pearson, W. 1990 Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **183**, 63–98. (doi:10.1016/0076-6879(90)83007-V)
- Pool, J. & Nielsen, R. 2007 Population size changes reshape genomic patterns of diversity. *Evolution* **61**, 3001–3006. (doi:10.1111/j.1558-5646.2007.00238.x)
- Pritchard, J. K., Stephens, M. & Donnelly, P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
- Ptak, S. E. & Przeworski, M. 2002 Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet.* **18**, 559–563. (doi:10.1016/S0168-9525(02)02781-6)
- Ray, N., Currat, M. & Excoffier, L. 2003 Intra-deme molecular diversity in spatially expanding populations. *Mol. Biol. Evol.* **20**, 76–86. (doi:10.1093/molbev/msg009)
- Rice, W. R. 1987 Genetic hitch-hiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**, 161–167.
- Swarbreck, D. *et al.* 2008 The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* **36**, D1009–D1014. (doi:10.1093/nar/gkm965)
- Tajima, F. 1989a Statistical method for testing the neutral mutation hypothesis. *Genetics* **123**, 585–595.
- Tajima, F. 1989b The effect of change in population size on DNA polymorphism. *Genetics* **123**, 597–601.
- Tajima, F. 1993 Statistical analysis of DNA polymorphism. *Jpn. J. Genet.* **68**, 567–595. (doi:10.1266/jgg.68.567)
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007 MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599. (doi:10.1093/molbev/msm092)
- Wakeley, J. & Aliacar, N. 2001 Gene genealogies in a metapopulation. *Genetics* **159**, 893–905.
- Wright, S. I. & Charlesworth, B. 2004 The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* **168**, 1071–1076. (doi:10.1534/genetics.104.026500)
- Yang, Z. 2001 Adaptive molecular evolution. In *Handbook of statistical genetics*, vol. 411 (eds D. J. Balding, M. Bishop & C. Cannings), pp. 327–350. New York, NY: John Wiley.
- Zeng, K., Fu, X. Y., Shi, S. & Wu, C. I. 2006 Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**, 1431–1439. (doi:10.1534/genetics.106.061432)